Chapter 4 - Random Variables

4.1 - Random Variables

Frequently, when an experiment is performed, we are interested mainly in some function of the outcome opposed to the actual outcome

↳ These real-valued functions defined on the sample space {⇒ random variables}

EXAMPLE 1b

Three balls are to be randomly selected without replacement from an urn containing 20 balls numbered 1 through 20. If we bet that at least one of the balls that are drawn has a number as large as or larger than 17, what is the probability that we win the bet?

Solution. Let X denote the largest number selected. Then X is a random variable taking on one of the values $3, 4, \ldots, 20$. Furthermore, if we suppose that each of the $\binom{20}{3}$ possible selections are equally likely to occur, then

$$P\{X = i\} = \frac{\binom{i-1}{2}}{\binom{20}{3}} \qquad i = 3, \ldots, 20 \tag{1.1}$$

Equation (1.1) follows because the number of selections that result in the event $\{X = i\}$ is just the number of selections that result in the ball numbered i and two of the balls numbered 1 through $i - 1$ being chosen. Because there are clearly $\binom{1}{1}\binom{i-1}{2}$ such selections, we obtain the probabilities expressed in Equation (1.1), from which we see that

$$P\{X = 20\} = \frac{\binom{19}{2}}{\binom{20}{3}} = \frac{3}{20} = .150$$

$$P\{X = 19\} = \frac{\binom{18}{2}}{\binom{20}{3}} = \frac{51}{380} \approx .134$$

$$P\{X = 18\} = \frac{\binom{17}{2}}{\binom{20}{3}} = \frac{34}{285} \approx .119$$

$$P\{X = 17\} = \frac{\binom{16}{2}}{\binom{20}{3}} = \frac{2}{19} \approx .105$$

Hence, since the event $\{X \geq 17\}$ is the union of the disjoint events $\{X = i\}$, $i = 17, 18, 19, 20$, it follows that the probability of our winning the bet is given by

$$P\{X \geq 17\} \approx .105 + .119 + .134 + .150 = .508 \qquad \blacksquare$$

More examples can be found on this chapter

4.2 - Discrete Random Variables

- A random variable that can take on at most a countable number of possible values is said to be discrete.
- For a discrete random variable X we define the probability mass function $p(a)$ of X by

$$p(a) = \mathbb{P}(X = a)$$

-> Probability mass function $p(a)$ is positive for at most a countable number of values of a.

That is, if X must assume one of the values $x_1, x_2, x_3 \ldots$; then

$p(x_i) \geq 0$ for $i = 1, 2, \ldots$

$p(x) = 0$ for all other values

Since X must take on <u>one</u> of the values x_i, we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

It is often instructive to present the probability mass function in a graphical format by plotting $p(x_i)$ on the y-axis against x_i on the x-axis. For instance, if the probability mass function of X is

$$p(0) = \frac{1}{4} \qquad p(1) = \frac{1}{2} \qquad p(2) = \frac{1}{4}$$

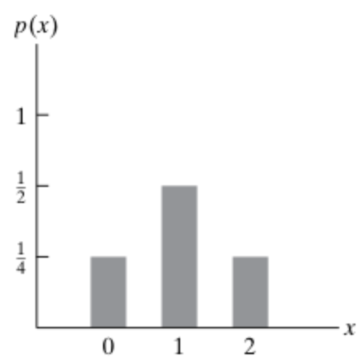

FIGURE 4.1

EXAMPLE 2a

The probability mass function of a random variable X is given by $p(i) = c\lambda^i/i!$, $i = 0, 1, 2, \ldots$, where λ is some positive value. Find (a) $P\{X = 0\}$ and (b) $P\{X > 2\}$.

Solution. Since $\sum_{i=0}^{\infty} p(i) = 1$, we have

$$c \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1$$

which, because $e^x = \sum_{i=0}^{\infty} x^i/i!$, implies that

$$ce^{\lambda} = 1 \quad \text{or} \quad c = e^{-\lambda}$$

Hence,

(a) $P\{X = 0\} = e^{-\lambda}\lambda^0/0! = e^{-\lambda}$

(b) $P\{X > 2\} = 1 - P\{X \le 2\} = 1 - P\{X = 0\} - P\{X = 1\} - P\{X = 2\}$

$$= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}$$ ■

The cumulative distribution function F can be expressed in terms of $p(a)$ by

$$F(a) = \sum_{\text{all } x \le a} p(x)$$

If X is a discrete random variable whose possible values are $x_1, x_2, x_3, \ldots$, where $x_1 < x_2 < x_3 < \cdots$, then the distribution function F of X is a step function. That is,

the value of F is constant in the intervals $[x_{i-1}, x_i)$ and then takes a step (or jump) of size $p(x_i)$ at x_i. For instance, if X has a probability mass function given by

$$p(1) = \frac{1}{4} \qquad p(2) = \frac{1}{2} \qquad p(3) = \frac{1}{8} \qquad p(4) = \frac{1}{8}$$

then its cumulative distribution function is

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{4} & 1 \le a < 2 \\ \frac{3}{4} & 2 \le a < 3 \\ \frac{7}{8} & 3 \le a < 4 \\ 1 & 4 \le a \end{cases}$$

This function is depicted graphically in Figure 4.3.

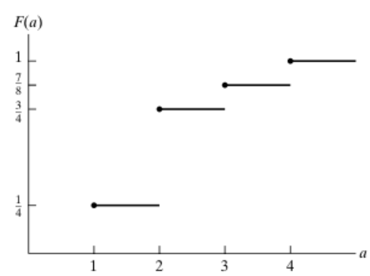

FIGURE 4.3

Note that the size of the step at any of the values 1, 2, 3, and 4 is equal to the probability that X assumes that particular value.

$\boxed{4.3 - \text{EXPECTED VALUE}}$ * SOS!

- One of the most important concepts in probability theory is that of expectation of a random variable
- If X is a discrete random variable having, a probability mass function $p(x)$, then the expectation / expected value of X, $E[X]$

$$\boxed{E[X] = \sum_{x : p(x) > 0} x p(x)}$$

In other words, the expected value of X is a weighted average of the possible values that X can take on. each value being weighted by the probability that X assumes.

For instance:

$$p(0) = \frac{1}{2} = p(1)$$

then

$$E[X] = 0\left(\frac{1}{2}\right) + 1\left(\frac{1}{2}\right) = \frac{1}{2}$$

is just the ordinary average of the two possible values, 0 and 1, that X can assume. On the other hand, if

$$p(0) = \frac{1}{3} \qquad p(1) = \frac{2}{3}$$

then

$$E[X] = 0\left(\frac{1}{3}\right) + 1\left(\frac{2}{3}\right) = \frac{2}{3}$$

is a weighted average of 2 possible values 0 & 1, where 1 is given twice as much weight. as the value 0., since $p(1) = 2p(0)$.

Note:

Another motivation of the definition of expectation is provided by the frequency interpretation of probabilities. This interpretation (partially justified by the strong law of large numbers, to be presented in Chapter 8) assumes that if an infinite sequence of independent replications of an experiment is performed, then, for any event E, the proportion of time that E occurs will be $P(E)$. Now, consider a random variable X that must take on one of the values $x_1, x_2, \ldots x_n$ with respective probabilities $p(x_1), p(x_2), \ldots, p(x_n)$, and think of X as representing our winnings in a single game of chance. That is, with probability $p(x_i)$ we shall win x_i units $i = 1, 2, \ldots, n$. By the frequency interpretation, if we play this game continually, then the proportion of time that we win x_i will be $p(x_i)$. Since this is true for all $i, i = 1, 2, \ldots, n$, it follows that our average winnings per game will be

Example.

$$\boxed{\sum_{i=1}^{n} x_i p(x_i) = E[X]}$$

EXAMPLE 3d

A school class of 120 students is driven in 3 buses to a symphonic performance. There are 36 students in one of the buses, 40 in another, and 44 in the third bus. When the buses arrive, one of the 120 students is randomly chosen. Let X denote the number of students on the bus of that randomly chosen student, and find $E[X]$.

Solution. Since the randomly chosen student is equally likely to be any of the 120 students, it follows that

$$P\{X = 36\} = \frac{36}{120} \quad P\{X = 40\} = \frac{40}{120} \quad P\{X = 44\} = \frac{44}{120}$$

Hence,

$$E[X] = 36\left(\frac{3}{10}\right) + 40\left(\frac{1}{3}\right) + 44\left(\frac{11}{30}\right) = \frac{1208}{30} = 40.2667$$

However, the average number of students on a bus is $120/3 = 40$, showing that the expected number of students on the bus of a randomly chosen student is larger than the average number of students on a bus. This is a general phenomenon, and it occurs because the more students there are on a bus, the more likely it is that a randomly chosen student would have been on that bus. As a result, buses with many students are given more weight than those with fewer students. (See Self-Test Problem 4.) ■

Remark. The probability concept of expectation is analogous to the physical concept of the *center of gravity* of a distribution of mass. Consider a discrete random variable X having probability mass function $p(x_i), i \geq 1$. If we now imagine a weightless rod in which weights with mass $p(x_i), i \geq 1$, are located at the points $x_i, i \geq 1$ (see Figure 4.4), then the point at which the rod would be in balance is known as the center of gravity. For those readers acquainted with elementary statics, it is now a simple matter to show that this point is at $E[X]$.† ■

$p(-1) = .10, \quad p(0) = .25, \quad p(1) = .30, \quad p(2) = .35$

^ = center of gravity = .9

FIGURE 4.4

4.4 - Expectation of a Function of a Random Variable

- Suppose that we are given a discrete random variable with its probability mass function and that we want to compute the expected value of some function of X, eg $g(x)$ (??)

↳ Since $g(x)$ is a discrete random variable, it has a probability mass function, which can be determined from the probability mass function of X.
Once we do that we can compute $E[g(X)]$ using definition.

EXAMPLE 4a

Let X denote a random variable that takes on any of the values -1, 0, and 1 with respective probabilities

$$P\{X = -1\} = .2 \quad P\{X = 0\} = .5 \quad P\{X = 1\} = .3$$

Compute $E[X^2]$.

Solution. Let $Y = X^2$. Then the probability mass function of Y is given by

$$P\{Y = 1\} = P\{X = -1\} + P\{X = 1\} = .5$$
$$P\{Y = 0\} = P\{X = 0\} = .5$$

Hence,

$$E[X^2] = E[Y] = 1(.5) + 0(.5) = .5$$

---

†To prove this, we must show that the sum of the torques tending to turn the point around $E[X]$ is equal to 0. That is, we must show that $0 = \sum_i (x_i - E[X])p(x_i)$, which is immediate.

Note that

$$.5 = E[X^2] \neq (E[X])^2 = .01 \qquad \blacksquare$$

Although the preceding procedure will always enable us to compute the expected value of any function of X from a knowledge of the probability mass function of X, there is another way of thinking about $E[g(X)]$: Since $g(X)$ will equal $g(x)$ whenever X is equal to x, it seems reasonable that $E[g(X)]$ should just be a weighted average of the values $g(x)$, with $g(x)$ being weighted by the probability that X is equal to x. That is, the following result is quite intuitive:

Proposition 4.1.
If X is a discrete random variable that takes on one of the values $x_i, i \geq 1$, with respective probabilities $p(x_i)$, then, for any real-valued function g,

$$E[g(X)] = \sum_i g(x_i)p(x_i)$$

Before proving this proposition, let us check that it is in accord with the results of Example 4a. Applying it to that example yields

$$\begin{aligned} E\{X^2\} &= (-1)^2(.2) + 0^2(.5) + 1^2(.3) \\ &= 1(.2 + .3) + 0(.5) \\ &= .5 \end{aligned}$$

which is in agreement with the result given in Example 4a.

Proof of Proposition 4.1: The proof of Proposition 4.1 proceeds, as in the preceding verification, by grouping together all the terms in $\sum_i g(x_i)p(x_i)$ having the same value of $g(x_i)$. Specifically, suppose that $y_j, j \geq 1$, represent the different values of $g(x_i), i \geq 1$. Then, grouping all the $g(x_i)$ having the same value gives

$$\begin{aligned} \sum_i g(x_i)p(x_i) &= \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i) \\ &= \sum_j \sum_{i:g(x_i)=y_j} y_j p(x_i) \\ &= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i) \\ &= \sum_j y_j P\{g(X) = y_j\} \\ &= E[g(X)] \end{aligned}$$

□

4.5 - VARIANCE

Definition

If X is a random variable with mean μ, then the variance of X, denoted by $\mathrm{Var}(X)$, is defined by

$$\mathrm{Var}(X) = E[(X - \mu)^2]$$

An alternative formula for $\mathrm{Var}(X)$ is derived as follows:

$$\begin{aligned}
\mathrm{Var}(X) &= E[(X - \mu)^2] \\
&= \sum_x (x - \mu)^2 p(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2) p(x) \\
&= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) \\
&= E[X^2] - 2\mu^2 + \mu^2 \\
&= E[X^2] - \mu^2
\end{aligned}$$

That is,

$$\boxed{\mathrm{Var}(X) = E[X^2] - (E[X])^2}$$

In words, the variance of X is equal to the expected value of X^2 minus the square of its expected value. In practice, this formula frequently offers the easiest way to compute $\mathrm{Var}(X)$.

EXAMPLE 5a

Calculate $\mathrm{Var}(X)$ if X represents the outcome when a fair die is rolled.

Solution. It was shown in Example 3a that $E[X] = \frac{7}{2}$. Also,

$$E[X^2] = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right)$$
$$= \left(\frac{1}{6}\right)(91)$$

Hence,

$$\mathrm{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \qquad \blacksquare$$

A useful identity is that, for any constants a and b,

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

To prove this equality, let $\mu = E[X]$ and note from Corollary 4.1 that $E[aX + b] = a\mu + b$. Therefore,

$$\begin{aligned}\mathrm{Var}(aX + b) &= E[(aX + b - a\mu - b)^2] \\ &= E[a^2(X - \mu)^2] \\ &= a^2E[(X - \mu)^2] \\ &= a^2\mathrm{Var}(X)\end{aligned}$$

Remarks. (a) Analogous to the means being the center of gravity of a distribution of mass, the variance represents, in the terminology of mechanics, the moment of inertia.

(b) The square root of the $\mathrm{Var}(X)$ is called the *standard deviation* of X, and we denote it by $\mathrm{SD}(X)$. That is,

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

Discrete random variables are often classified according to their probability mass functions. In the next few sections, we consider some of the more common types.

4.6 - BERNOULLI & BINOMIAL RANDOM VARIABLES

Suppose a trial or experiment whose outcome can be either success or failure.

If we let $X = 1$, when outcome is a success and $X = 0$ when failure.

Then probability mass function of X

$$\boxed{\left.\begin{array}{l} p(0) = \mathbb{P}(X=0) = 1-p \\ p(1) = \mathbb{P}(X=1) = p. \end{array}\right\} \; 0 \leq p \leq 1} \quad 6.1$$

$\Rightarrow$ A random variable X is said to be Bernoulli random variable. if its probability mass function is given be 6.1 for some $p \in (0,1)$

Note:

Suppose now than n independent trials, each of which results in a success with probability (p) & failure with probability $(1-p)$.
If X represents the number of successes that occur in n trials $\Leftrightarrow$ X is said to be binomial random variable. with parameters (n, p).
$\therefore$ Bernoulli random variable is just a binomial random variable with parameters $(1, p)$.

The probability mass function of binomial random variable parameters (n, p)

$\Rightarrow p(i) = \binom{n}{i} p^i (1-p)^{n-i}$

$i = 0, 1, 2, 3, \ldots, n$ 6.2

Validating equation (6.2)

Noting that the probability of any particular sequence of (n) outcomes containing (i) successes and ($n-i$) failures is by the assumed independence of trials

$p^i (1-p)^{n-i}$

$\Rightarrow$ equation 6.2 follows since

there are $\binom{n}{i}$ different sequences of the n outcomes

→ leading to $\binom{n}{i}$ successes and $(n-i)$ failures

eg

$n=4, i=2 \Rightarrow \binom{4}{2}=6$

in which the 4 trials can result into 2 successes.

Note that, by the binomial theorem, the probabilities sum to 1; that is,

$$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = [p + (1-p)]^n = 1$$

EXAMPLE 6b

It is known that screws produced by a certain company will be defective with probability .01, independently of each other. The company sells the screws in packages of 10 and offers a money-back guarantee that at most 1 of the 10 screws is defective. What proportion of packages sold must the company replace?

Solution. If X is the number of defective screws in a package, then X is a binomial random variable with parameters (10, .01). Hence, the probability that a package will have to be replaced is

$$1 - P\{X=0\} - P\{X=1\} = 1 - \binom{10}{0}(.01)^0(.99)^{10} - \binom{10}{1}(.01)^1(.99)^9 \approx .004$$

Thus, only .4 percent of the packages will have to be replaced. ■

4.6.1 Properties of Binomial Random Variables

We will now examine the properties of a binomial random variable with parameters n and p. To begin, let us compute its expected value and variance. Now,

$$\begin{aligned} E[X^k] &= \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=1}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} \end{aligned}$$

Using the identity

$$i\binom{n}{i} = n\binom{n-1}{i-1}$$

gives

$$\begin{aligned} E[X^k] &= np \sum_{i=1}^{n} i^{k-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \\ &= np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \quad \begin{array}{l}\text{by letting} \\ j = i-1\end{array} \\ &= npE[(Y+1)^{k-1}] \end{aligned}$$

where Y is a binomial random variable with parameters $n-1, p$. Setting $k = 1$ in the preceding equation yields

$$E[X] = np$$

That is, the expected number of successes that occur in n independent trials when each is a success with probability p is equal to np. Setting $k = 2$ in the preceding equation, and using the preceding formula for the expected value of a binomial random variable yields

$$\begin{aligned} E[X^2] &= npE[Y+1] \\ &= np[(n-1)p+1] \end{aligned}$$

Since $E[X] = np$, we obtain

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= np[(n-1)p+1] - (np)^2 \\ &= np(1-p) \end{aligned}$$

Summing up, we have shown the following:

If X is a binomial random variable with parameters n and p, then

$$\begin{aligned} E[X] &= np \\ \text{Var}(X) &= np(1-p) \end{aligned}$$

The following proposition details how the binomial probability mass function first increases and then decreases.

Proposition 6.1. If X is a binomial random variable with parameters (n, p), where $0 < p < 1$, then as k goes from 0 to n, $P\{X = k\}$ first increases monotonically and then decreases monotonically, reaching its largest value when k is the largest integer less than or equal to $(n + 1)p$.

Proof. We prove the proposition by considering $P\{X = k\}/P\{X = k - 1\}$ and determining for what values of k it is greater or less than 1. Now,

$$
\begin{aligned}
\frac{P\{X=k\}}{P\{X=k-1\}} &= \frac{\dfrac{n!}{(n-k)!k!}p^k(1-p)^{n-k}}{\dfrac{n!}{(n-k+1)!(k-1)!}p^{k-1}(1-p)^{n-k+1}} \\
&= \frac{(n-k+1)p}{k(1-p)}
\end{aligned}
$$

Hence, $P\{X = k\} \geq P\{X = k - 1\}$ if and only if

$$(n - k + 1)p \geq k(1 - p)$$

or, equivalently, if and only if

$$k \leq (n + 1)p$$

and the proposition is proved. ■

As an illustration of Proposition 6.1 consider Figure 4.5, the graph of the probability mass function of a binomial random variable with parameters $(10, \frac{1}{2})$.

EXAMPLE 6g

In a U.S. presidential election, the candidate who gains the maximum number of votes in a state is awarded the total number of electoral college votes allocated to

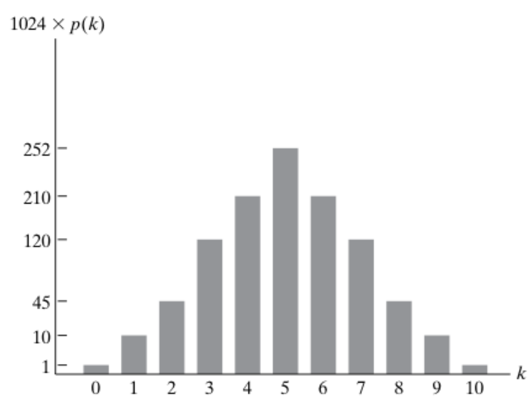

FIGURE 4.5 Graph of $p(k) = \binom{10}{k}\left(\frac{1}{2}\right)^{10}$

that state. The number of electoral college votes of a given state is roughly proportional to the population of that state—that is, a state with population n has roughly nc electoral votes. (Actually, it is closer to $nc + 2$, as a state is given an electoral vote for each member it has in the House of Representatives, with the number of such representatives being roughly proportional to the population of the state, and one electoral college vote for each of its two senators.) Let us determine the average power of a citizen in a state of size n in a close presidential election, where, by *average power in a close election*, we mean that a voter in a state of size $n = 2k + 1$ will be decisive if the other $n - 1$ voters split their votes evenly between the two candidates. (We are assuming here that n is odd, but the case where n is even is quite similar.) Because the election is close, we shall suppose that each of the other $n - 1 = 2k$ voters acts independently and is equally likely to vote for either candidate. Hence, the probability that a voter in a state of size $n = 2k + 1$ will make a difference to the outcome is the same as the probability that $2k$ tosses of a fair coin land heads and tails an equal number of times. That is,

$$\begin{aligned}
&P\{\text{voter in state of size } 2k + 1 \text{ makes a difference}\} \\
&\qquad = \binom{2k}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^k \\
&\qquad = \frac{(2k)!}{k!k!2^{2k}}
\end{aligned}$$

To approximate the preceding equality, we make use of Stirling's approximation, which says that, for k large,

$$k! \sim k^{k+1/2}e^{-k}\sqrt{2\pi}$$

where we say that $a_k \sim b_k$ when the ratio a_k/b_k approaches 1 as k approaches ∞. Hence, it follows that

$$\begin{aligned}
&P\{\text{voter in state of size } 2k + 1 \text{ makes a difference}\} \\
&\qquad \sim \frac{(2k)^{2k+1/2}e^{-2k}\sqrt{2\pi}}{k^{2k+1}e^{-2k}(2\pi)2^{2k}} = \frac{1}{\sqrt{k\pi}}
\end{aligned}$$

Because such a voter (if he or she makes a difference) will affect nc electoral votes, the expected number of electoral votes a voter in a state of size n will affect—or the voter's average power—is given by

$$\begin{aligned}
\text{average power} &= ncP\{\text{makes a difference}\} \\
&\sim \frac{nc}{\sqrt{n\pi/2}} \\
&= c\sqrt{2n/\pi}
\end{aligned}$$

Thus, the average power of a voter in a state of size n is proportional to the square root of n, showing that, in presidential elections, voters in large states have more power than do those in smaller states. ■

4.6.2 Computing the Binomial Distribution Function

Suppose that X is binomial with parameters (n, p). The key to computing its distribution function

$$P\{X \le i\} = \sum_{k=0}^{i} \binom{n}{k} p^k (1 - p)^{n-k} \qquad i = 0, 1, \ldots, n$$

is to utilize the following relationship between $P\{X = k + 1\}$ and $P\{X = k\}$, which was established in the proof of Proposition 6.1:

$$P\{X = k + 1\} = \frac{p}{1 - p}\frac{n - k}{k + 1}P\{X = k\} \tag{6.3}$$

EXAMPLE 6h

Let X be a binomial random variable with parameters $n = 6, p = .4$. Then, starting with $P\{X = 0\} = (.6)^6$ and recursively employing Equation (6.3), we obtain

$$\begin{aligned}
P\{X = 0\} &= (.6)^6 \approx .0467 \\
P\{X = 1\} &= \frac{4}{6}\frac{6}{1}P\{X = 0\} \approx .1866 \\
P\{X = 2\} &= \frac{4}{6}\frac{5}{2}P\{X = 1\} \approx .3110 \\
P\{X = 3\} &= \frac{4}{6}\frac{4}{3}P\{X = 2\} \approx .2765 \\
P\{X = 4\} &= \frac{4}{6}\frac{3}{4}P\{X = 3\} \approx .1382 \\
P\{X = 5\} &= \frac{4}{6}\frac{2}{5}P\{X = 4\} \approx .0369 \\
P\{X = 6\} &= \frac{4}{6}\frac{1}{6}P\{X = 5\} \approx .0041
\end{aligned}$$

■

A computer program that utilizes the recursion (6.3) to compute the binomial distribution function is easily written. To compute $P\{X \le i\}$, the program should first compute $P\{X = i\}$ and then use the recursion to successively compute $P\{X = i - 1\}, P\{X = i - 2\}$, and so on.

4.7 - POISSON RANDOM VARIABLE

A random variable X that takes on one of the values $0, 1, 2, \ldots$ is said to be a Poisson random variable with parameters λ.

$$p(i) = P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!} \quad i = 0, 1, 2, 3 \ldots$$

Equation (7.1) defines a probability mass function, since

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1$$

* **Note** *.

Poisson can be used as an approximation for binomial random variable with parameters (n, p) when n is large &

p is small

size. To see this, suppose that X is a binomial random variable with parameters (n, p), and let $\lambda = np$. Then

$$\begin{aligned}
P\{X = i\} &= \frac{n!}{(n-i)!\,i!}p^i(1-p)^{n-i} \\
&= \frac{n!}{(n-i)!\,i!}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i} \\
&= \frac{n(n-1)\cdots(n-i+1)}{n^i}\frac{\lambda^i}{i!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}
\end{aligned}$$

Now, for n large and λ moderate,

$$\left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda} \quad \frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1 \quad \left(1-\frac{\lambda}{n}\right)^i \approx 1$$

Hence, for n large and λ moderate,

$$P\{X = i\} \approx e^{-\lambda}\frac{\lambda^i}{i!}$$

EXAMPLE 7b

Suppose that the probability that an item produced by a certain machine will be defective is .1. Find the probability that a sample of 10 items will contain at most 1 defective item.

Solution. The desired probability is $\binom{10}{0}(.1)^0(.9)^{10} + \binom{10}{1}(.1)^1(.9)^9 = .7361$, whereas the Poisson approximation yields the value $e^{-1} + e^{-1} \approx .7358$. ■

* EXAMPLE 7C pg 160 *
explanation + Example 7d

For a second illustration of the strength of the Poisson approximation when the trials are weakly dependent, let us consider again the birthday problem presented in Example 5i of Chapter 2. In this example, we suppose that each of n people is equally likely to have any of the 365 days of the year as his or her birthday, and the problem

is to determine the probability that a set of n independent people all have different birthdays. A combinatorial argument was used to determine this probability, which was shown to be less than $\frac{1}{2}$ when $n = 23$.

We can approximate the preceding probability by using the Poisson approximation as follows: Imagine that we have a trial for each of the $\binom{n}{2}$ pairs of individuals i and $j, i \neq j$, and say that trial i, j is a success if persons i and j have the same birthday. If we let E_{ij} denote the event that trial i, j is a success, then, whereas the $\binom{n}{2}$ events $E_{ij}, 1 \leq i < j \leq n$, are not independent (see Theoretical Exercise 21), their dependence appears to be rather weak. (Indeed, these events are even *pairwise independent*, in that any 2 of the events E_{ij} and E_{kl} are independent—again, see Theoretical Exercise 21). Since $P(E_{ij}) = 1/365$, it is reasonable to suppose that the number of successes should approximately have a Poisson distribution with mean $\binom{n}{2}\Big/365 = n(n-1)/730$. Therefore,

$$
\begin{aligned}
P\{\text{no 2 people have the same birthday}\} &= P\{0 \text{ successes}\} \\
&\approx \exp\left\{\frac{-n(n-1)}{730}\right\}
\end{aligned}
$$

To determine the smallest integer n for which this probability is less than $\frac{1}{2}$, note that

$$
\exp\left\{\frac{-n(n-1)}{730}\right\} \leq \frac{1}{2}
$$

is equivalent to

$$
\exp\left\{\frac{n(n-1)}{730}\right\} \geq 2
$$

Taking logarithms of both sides, we obtain

$$
\begin{aligned}
n(n-1) &\geq 730 \log 2 \\
&\approx 505.997
\end{aligned}
$$

which yields the solution $n = 23$, in agreement with the result of Example 5i of Chapter 2.

Suppose now that we wanted the probability that, among the n people, no 3 of them have their birthday on the same day. Whereas this now becomes a difficult combinatorial problem, it is a simple matter to obtain a good approximation. To begin, imagine that we have a trial for each of the $\binom{n}{3}$ triplets i, j, k, where $1 \leq i < j < k \leq n$, and call the i, j, k trial a success if persons i, j, and k all have their birthday on the same day. As before, we can then conclude that the number of successes is approximately a Poisson random variable with parameter

$$
\begin{aligned}
\binom{n}{3} P\{i, j, k \text{ have the same birthday}\} &= \binom{n}{3}\left(\frac{1}{365}\right)^2 \\
&= \frac{n(n-1)(n-2)}{6 \times (365)^2}
\end{aligned}
$$

Hence,

$$
P\{\text{no 3 have the same birthday}\} \approx \exp\left\{\frac{-n(n-1)(n-2)}{799350}\right\}
$$

This probability will be less than $\frac{1}{2}$ when n is such that

$$
n(n-1)(n-2) \geq 799350 \log 2 \approx 554067.1
$$

which is equivalent to $n \geq 84$. Thus, the approximate probability that at least 3 people in a group of size 84 or larger will have the same birthday exceeds $\frac{1}{2}$.

For the number of events to occur to approximately have a Poisson distribution, it is not essential that all the events have the same probability of occurrence, but only that all of these probabilities be small. The following is referred to as the *Poisson paradigm*.

Poisson Paradigm. Consider n events, with p_i equal to the probability that event i occurs, $i = 1, \ldots, n$. If all the p_i are "small" and the trials are either independent or at most "weakly dependent," then the number of these events that occur approximately has a Poisson distribution with mean $\sum_{i=1}^{n} p_i$.

Our next example not only makes use of the Poisson paradigm, but also illustrates a variety of the techniques we have studied so far.

4.8 - OTHER DISCRETE PROBABILITY DISTRIBUTIONS

- 4.8.1 - The Geometric Random Variable

- Suppose, independent trials, each having a probability p, $0 < p < 1$, of being a success, are performed until a success occurs.
If we let X equal the number of trials required then:

$$\boxed{\mathbb{P}(X = n) = (1-p)^{n-1}\, p \qquad n = 1, 2, \ldots} \qquad (8.1)$$

① In order for X to equal n, it is necessary & sufficient that the first $(n-1)$ trials are failures & the n^{th} trial is a success.

② Trials are assumed to be independent since:

$$\sum_{n=1}^{\infty} P\{X = n\} = p \sum_{n=1}^{\infty} (1 - p)^{n-1} = \frac{p}{1 - (1 - p)} = 1$$

* Any random variable X whose probability mass function is given by equation 8.1, is said to be a geometric random variable with parameter p.

EXAMPLE 8b

Find the expected value of a geometric random variable.

Solution. With $q = 1 - p$, we have

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{\infty} i q^{i-1} p \\
&= \sum_{i=1}^{\infty} (i - 1 + 1) q^{i-1} p \\
&= \sum_{i=1}^{\infty} (i - 1) q^{i-1} p + \sum_{i=1}^{\infty} q^{i-1} p \\
&= \sum_{j=0}^{\infty} j q^{j} p + 1 \\
&= q \sum_{j=1}^{\infty} j q^{j-1} p + 1 \\
&= q E[X] + 1
\end{aligned}
$$

Hence,

$$pE[X] = 1$$

yielding the result

$$E[X] = \frac{1}{p}$$

In other words, if independent trials having a common probability p of being successful are performed until the first success occurs, then the expected number of required trials equals $1/p$. For instance, the expected number of rolls of a fair die that it takes to obtain the value 1 is 6. ■

EXAMPLE 8c

Find the variance of a geometric random variable.

Solution. To determine $\mathrm{Var}(X)$, let us first compute $E[X^2]$. With $q = 1 - p$, we have

$$\begin{aligned}
E[X^2] &= \sum_{i=1}^{\infty} i^2 q^{i-1} p \\
&= \sum_{i=1}^{\infty} (i - 1 + 1)^2 q^{i-1} p \\
&= \sum_{i=1}^{\infty} (i - 1)^2 q^{i-1} p + \sum_{i=1}^{\infty} 2(i - 1) q^{i-1} p + \sum_{i=1}^{\infty} q^{i-1} p \\
&= \sum_{j=0}^{\infty} j^2 q^j p + 2 \sum_{j=1}^{\infty} j q^j p + 1 \\
&= qE[X^2] + 2qE[X] + 1
\end{aligned}$$

Using $E[X] = 1/p$, the equation for $E[X^2]$ yields

$$pE[X^2] = \frac{2q}{p} + 1$$

Hence,

$$E[X^2] = \frac{2q + p}{p^2} = \frac{q + 1}{p^2}$$

giving the result

$$\mathrm{Var}(X) = \frac{q + 1}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} = \frac{1 - p}{p^2}$$

■

4.8.2 - The Negative Binomial Random Variable

Suppose that independent trials, each having probability p, $0<p<1$, of being a success are performed, until a total of r successes is accumulated.

If we let X equal the number of trials required then,

$$P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r} \quad n = r, r+1, \ldots \tag{8.2}$$

↳ Follows, because in order for the r^{th} success to occur at the n^{th} trial, there must be $r-1$ successes in the first $n-1$ trials & n^{th} trial must be a success.

- The probability of the first event

$$\binom{n-1}{r-1} p^{r-1}(1-p)^{n-r}$$

& the probability of the second is p.

∴ independent.

→ To verify that a total of r successes must eventually be accumulated.

we can prove analytically that

$$\sum_{n=r}^{\infty} P\{X = n\} = \sum_{n=r}^{\infty} \binom{n-1}{r-1} p^{r}(1-p)^{n-r} = 1 \tag{8.3}$$

or we can give a probabilistic argument as follows: The number of trials required to obtain r successes can be expressed as $Y_1 + Y_2 + \cdots + Y_r$, where Y_1 equals the number of trials required for the first success, Y_2 the number of additional trials after the first success until the second success occurs, Y_3 the number of additional trials until the third success, and so on. Because the trials are independent and all have the same probability of success, it follows that $Y_1, Y_2, \ldots, Y_r$ are all geometric random variables. Hence, each is finite with probability 1, so $\sum_{i=1}^{r} Y_i$ must also be finite, establishing Equation (8.3).

Any random variable X whose probability mass function is given by Equation (8.2) is said to be a *negative binomial* random variable with parameters (r, p). Note that a geometric random variable is just a negative binomial with parameter $(1, p)$.

In the next example, we use the negative binomial to obtain another solution of the problem of the points.

EXAMPLE 8e ***The Banach match problem***

At all times, a pipe-smoking mathematician carries 2 matchboxes—1 in his left-hand pocket and 1 in his right-hand pocket. Each time he needs a match, he is equally likely to take it from either pocket. Consider the moment when the mathematician first discovers that one of his matchboxes is empty. If it is assumed that both matchboxes initially contained N matches, what is the probability that there are exactly k matches, $k = 0, 1, \ldots, N$, in the other box?

Solution. Let E denote the event that the mathematician first discovers that the right-hand matchbox is empty and that there are k matches in the left-hand box at the time. Now, this event will occur if and only if the $(N + 1)$th choice of the right-hand matchbox is made at the $(N + 1 + N - k)$th trial. Hence, from Equation (8.2) (with $p = \frac{1}{2}, r = N + 1$, and $n = 2N - k + 1$), we see that

$$P(E) = \binom{2N - k}{N}\left(\frac{1}{2}\right)^{2N-k+1}$$

Since there is an equal probability that it is the left-hand box that is first discovered to be empty and there are k matches in the right-hand box at that time, the desired result is

$$2P(E) = \binom{2N - k}{N}\left(\frac{1}{2}\right)^{2N-k}$$

■

EXAMPLE 8g

Find the expected value and the variance of the number of times one must throw a die until the outcome 1 has occurred 4 times.

Solution. Since the random variable of interest is a negative binomial with parameters $r = 4$ and $p = \frac{1}{6}$, it follows that

$$E[X] = 24$$

$$\text{Var}(X) = \frac{4\left(\frac{5}{6}\right)}{\left(\frac{1}{6}\right)^2} = 120$$

■

EXAMPLE 8f

Compute the expected value and the variance of a negative binomial random variable with parameters r and p.

Solution. We have

$$
\begin{aligned}
E[X^k] &= \sum_{n=r}^{\infty} n^k \binom{n-1}{r-1} p^r (1-p)^{n-r} \\
&= \frac{r}{p} \sum_{n=r}^{\infty} n^{k-1} \binom{n}{r} p^{r+1} (1-p)^{n-r} \quad \text{since} \quad n\binom{n-1}{r-1} = r\binom{n}{r} \\
&= \frac{r}{p} \sum_{m=r+1}^{\infty} (m-1)^{k-1} \binom{m-1}{r} p^{r+1} (1-p)^{m-(r+1)} \quad \begin{array}{c}\text{by setting} \\ m = n+1\end{array} \\
&= \frac{r}{p} E[(Y-1)^{k-1}]
\end{aligned}
$$

where Y is a negative binomial random variable with parameters $r+1, p$. Setting $k=1$ in the preceding equation yields

$$E[X] = \frac{r}{p}$$

Setting $k=2$ in the equation for $E[X^k]$ and using the formula for the expected value of a negative binomial random variable gives

$$
\begin{aligned}
E[X^2] &= \frac{r}{p} E[Y-1] \\
&= \frac{r}{p}\left(\frac{r+1}{p} - 1\right)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathrm{Var}(X) &= \frac{r}{p}\left(\frac{r+1}{p} - 1\right) - \left(\frac{r}{p}\right)^2 \\
&= \frac{r(1-p)}{p^2}
\end{aligned}
$$

■

Thus, from Example 8f, if independent trials, each of which is a success with probability p, are performed, then the expected value and variance of the number of trials that it takes to amass r successes is r/p and $r(1-p)/p^2$, respectively.

Since a geometric random variable is just a negative binomial with parameter $r=1$, it follows from the preceding example that the variance of a geometric random variable with parameter p is equal to $(1-p)/p^2$, which checks with the result of Example 8c.

4.8.3- The Hypergeometric Random Variable

Suppose that a sample of size n is to be chosen randomly (without replacement) from an urn containing N balls, of which m are white, & $N-m$ are black. If we let X denote the number of white balls selected, then

$$P\{X = i\} = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \qquad i = 0, 1, \ldots, n \qquad (8.4)$$

↳ A random Variable X, whose probability mass function is given by (8.4) for some values on n, N, m

↳ Hypergeometric random variable

Remark. Although we have written the hypergeometric probability mass function with i going from 0 to n, $P\{X = i\}$ will actually be 0, unless i satisfies the inequalities $n - (N - m) \le i \le \min(n, m)$. However, Equation (8.4) is always valid because of our convention that $\binom{r}{k}$ is equal to 0 when either $k < 0$ or $r < k$. ■

Note :

EXAMPLE 8h

An unknown number, say, N, of animals inhabit a certain region. To obtain some information about the size of the population, ecologists often perform the following experiment: They first catch a number, say, m, of these animals, mark them in some manner, and release them. After allowing the marked animals time to disperse throughout the region, a new catch of size, say, n, is made. Let X denote the number of marked animals in this second capture. If we assume that the population of animals in the region remained fixed between the time of the two catches and that each time an animal was caught it was equally likely to be any of the remaining uncaught animals, it follows that X is a hypergeometric random variable such that

$$P\{X = i\} = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \equiv P_i(N)$$

Suppose now that X is observed to equal i. Then, since $P_i(N)$ represents the probability of the observed event when there are actually N animals present in the region, it would appear that a reasonable estimate of N would be the value of N that maximizes $P_i(N)$. Such an estimate is called a *maximum likelihood* estimate. (See Theoretical Exercises 13 and 18 for other examples of this type of estimation procedure.)

The maximization of $P_i(N)$ can be done most simply by first noting that

$$\frac{P_i(N)}{P_i(N-1)} = \frac{(N-m)(N-n)}{N(N-m-n+i)}$$

Now, the preceding ratio is greater than 1 if and only if

$$(N-m)(N-n) \ge N(N-m-n+i)$$

or, equivalently, if and only if

$$N \le \frac{mn}{i}$$

Thus, $P_i(N)$ is first increasing and then decreasing, and reaches its maximum value at the largest integral value not exceeding mn/i. This value is the maximum likelihood estimate of N. For example, suppose that the initial catch consisted of $m = 50$ animals, which are marked and then released. If a subsequent catch consists of $n = 40$ animals of which $i = 4$ are marked, then we would estimate that there are some 500 animals in the region. (Note that the preceding estimate could also have been obtained by assuming that the proportion of marked animals in the region, m/N, is approximately equal to the proportion of marked animals in our second catch, i/n.) ■

EXAMPLE 8j

Determine the expected value and the variance of X, a hypergeometric random variable with parameters n, N, and m.

Solution.

$$\begin{aligned} E[X^k] &= \sum_{i=0}^{n} i^k P\{X = i\} \\ &= \sum_{i=1}^{n} i^k \binom{m}{i}\binom{N-m}{n-i} \Big/ \binom{N}{n} \end{aligned}$$

Using the identities

$$i\binom{m}{i} = m\binom{m-1}{i-1} \quad \text{and} \quad n\binom{N}{n} = N\binom{N-1}{n-1}$$

we obtain

$$\begin{aligned} E[X^k] &= \frac{nm}{N}\sum_{i=1}^{n} i^{k-1}\binom{m-1}{i-1}\binom{N-m}{n-i}\Big/\binom{N-1}{n-1} \\ &= \frac{nm}{N}\sum_{j=0}^{n-1}(j+1)^{k-1}\binom{m-1}{j}\binom{N-m}{n-1-j}\Big/\binom{N-1}{n-1} \\ &= \frac{nm}{N}E[(Y+1)^{k-1}] \end{aligned}$$

where Y is a hypergeometric random variable with parameters $n - 1$, $N - 1$, and $m - 1$. Hence, upon setting $k = 1$, we have

$$E[X] = \frac{nm}{N}$$

In words, if n balls are randomly selected from a set of N balls, of which m are white, then the expected number of white balls selected is nm/N.

Upon setting $k = 2$ in the equation for $E[X^k]$, we obtain

$$\begin{aligned} E[X^2] &= \frac{nm}{N}E[Y+1] \\ &= \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1\right] \end{aligned}$$

where the final equality uses our preceding result to compute the expected value of the hypergeometric random variable Y.

Because $E[X] = nm/N$, we can conclude that

$$\text{Var}(X) = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right]$$

Letting $p = m/N$ and using the identity

$$\frac{m-1}{N-1} = \frac{Np-1}{N-1} = p - \frac{1-p}{N-1}$$

shows that

$$\begin{aligned}\mathrm{Var}(X) &= np[(n-1)p - (n-1)\frac{1-p}{N-1} + 1 - np] \\ &= np(1-p)(1 - \frac{n-1}{N-1})\end{aligned}$$

■

Remark. We have shown in Example 8j that if n balls are randomly selected without replacement from a set of N balls, of which the fraction p are white, then the expected number of white balls chosen is np. In addition, if N is large in relation to n [so that $(N-n)/(N-1)$ is approximately equal to 1], then

$$\mathrm{Var}(X) \approx np(1-p)$$

In other words, $E[X]$ is the same as when the selection of the balls is done with replacement (so that the number of white balls is binomial with parameters n and p), and if the total collection of balls is large, then $\mathrm{Var}(X)$ is approximately equal to what it would be if the selection were done with replacement. This is, of course, exactly what we would have guessed, given our earlier result that when the number of balls in the urn is large, the number of white balls chosen approximately has the mass function of a binomial random variable. ■

4.8.4 - The Zeta (or Zipf) Distribution

A random variable is said to have a zeta distribution if probability mass function is given by

$$P\{X = k\} = \frac{C}{k^{\alpha+1}} \quad k = 1, 2, \ldots$$

for some value of $\alpha > 0$. Since the sum of the foregoing probabilities must equal 1, it follows that

$$C = \left[\sum_{k=1}^{\infty} \left(\frac{1}{k} \right)^{\alpha+1} \right]^{-1}$$

The zeta distribution owes its name to the fact that the function

$$\zeta(s) = 1 + \left(\frac{1}{2}\right)^s + \left(\frac{1}{3}\right)^s + \cdots + \left(\frac{1}{k}\right)^s + \cdots$$

is known in mathematical disciplines as the Riemann zeta function (after the German mathematician G. F. B. Riemann).

The zeta distribution was used by the Italian economist V. Pareto to describe the distribution of family incomes in a given country. However, it was G. K. Zipf who applied zeta distribution to a wide variety of problems in different areas and, in doing so, popularized its use.

4.9 - EXPECTED VALUE OF SUMS OF RANDOM VARIABLES

- A very important property of expectations is that the expected value of a sum of random variables is equal to the sum of their expectations

In this section we will prove this result under the assumption that the set of possible values of the probability experiment → sample space S

finite

countably infinite

So, for the remainder of this section, suppose that the sample space S is either a finite or a countably infinite set.

For a random variable X, let $X(s)$ denote the value of X when $s \in S$ is the outcome of the experiment. Now, if X and Y are both random variables, then so is their sum. That is, $Z = X + Y$ is also a random variable. Moreover, $Z(s) = X(s) + Y(s)$.

EXAMPLE 9a Important.

Suppose that the experiment consists of flipping a coin 5 times, with the outcome being the resulting sequence of heads and tails. Suppose X is the number of heads in the first 3 flips and Y is the number of heads in the final 2 flips. Let $Z = X + Y$. Then, for instance, for the outcome $s = (h, t, h, t, h)$,

$$X(s) = 2$$
$$Y(s) = 1$$
$$Z(s) = X(s) + Y(s) = 3$$

meaning that the outcome (h, t, h, t, h) results in 2 heads in the first three flips, 1 head in the final two flips, and a total of 3 heads in the five flips. ■

Let $p(s) = P(\{s\})$ be the probability that s is the outcome of the experiment. Because we can write any event A as the finite or countably infinite union of the mutually exclusive events $\{s\}, s \in A$, it follows by the axioms of probability that

$$P(A) = \sum_{s \in A} p(s)$$

When $A = S$, the preceding equation gives

$$1 = \sum_{s \in S} p(s)$$

Now, let X be a random variable, and consider $E[X]$. Because $X(s)$ is the value of X when s is the outcome of the experiment, it seems intuitive that $E[X]$—the weighted average of the possible values of X, with each value weighted by the probability that X assumes that value—should equal a weighted average of the values $X(s), s \in S$, with $X(s)$ weighted by the probability that s is the outcome of the experiment. We now prove this intuition.

Proposition 9.1.

$$E[X] = \sum_{s \in S} X(s)\,p(s)$$

Proof. Suppose that the distinct values of X are $x_i, i \geq 1$. For each i, let S_i be the event that X is equal to x_i. That is, $S_i = \{s : X(s) = x_i\}$. Then,

$$\begin{aligned}
E[X] &= \sum_i x_i P\{X = x_i\} \\
&= \sum_i x_i P(S_i) \\
&= \sum_i x_i \sum_{s \in S_i} p(s) \\
&= \sum_i \sum_{s \in S_i} x_i p(s) \\
&= \sum_i \sum_{s \in S_i} X(s) p(s) \\
&= \sum X(s) p(s)
\end{aligned}$$

EXAMPLE 9b

Suppose that two independent flips of a coin that comes up heads with probability p are made, and let X denote the number of heads obtained. Because

$$\begin{aligned}
P(X = 0) &= P(t,t) = (1 - p)^2, \\
P(X = 1) &= P(h,t) + P(t,h) = 2p(1 - p) \\
P(X = 2) &= P(h,h) = p^2
\end{aligned}$$

it follows from the definition of expected value that

$$E[X] = 0 \cdot (1 - p)^2 + 1 \cdot 2p(1 - p) + 2 \cdot p^2 = 2p$$

which agrees with

$$\begin{aligned}
E[X] &= X(h,h)p^2 + X(h,t)p(1 - p) + X(t,h)(1 - p)p + X(t,t)(1 - p)^2 \\
&= 2p^2 + p(1 - p) + (1 - p)p \\
&= 2p
\end{aligned}$$

■

We now prove the important and useful result that the expected value of a sum of random variables is equal to the sum of their expectations.

Corollary 9.2. For random variables $X_1, X_2, \ldots, X_n$,

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Proof. Let $Z = \sum_{i=1}^{n} X_i$. Then, by Proposition 9.1,

$$\begin{aligned}
E[Z] &= \sum_{s \in S} Z(s)p(s) \\
&= \sum_{s \in S} \big(X_1(s) + X_2(s) + \ldots + X_n(s)\big)\, p(s) \\
&= \sum_{s \in S} X_1(s)p(s) + \sum_{s \in S} X_2(s)p(s) + \ldots + \sum_{s \in S} X_n(s)p(s) \\
&= E[X_1] + E[X_2] + \ldots + E[X_n]
\end{aligned}$$

■

4.10 – Properties of the cumulative Distribution functions

Recall that, for the distribution function F of X, $F(b)$ denotes the probability that the random variable X takes on a value that is less than or equal to b. Following are some properties of the cumulative distribution function (c.d.f.) F:

1. F is a nondecreasing function; that is, if $a < b$, then $F(a) \leq F(b)$.
2. $\lim_{b \to \infty} F(b) = 1$.
3. $\lim_{b \to -\infty} F(b) = 0$.
4. F is right continuous. That is, for any b and any decreasing sequence $b_n, n \geq 1$, that converges to b, $\lim_{n \to \infty} F(b_n) = F(b)$.

EXAMPLE 10a

The distribution function of the random variable X is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{2} & 0 \le x < 1 \\ \dfrac{2}{3} & 1 \le x < 2 \\ \dfrac{11}{12} & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

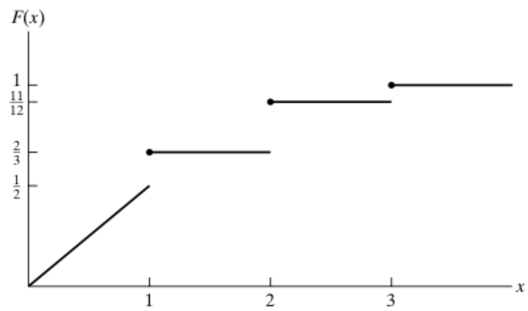

FIGURE 4.8: Graph of $F(x)$.

A graph of $F(x)$ is presented in Figure 4.8. Compute (a) $P\{X < 3\}$, (b) $P\{X = 1\}$, (c) $P\{X > \frac{1}{2}\}$, and (d) $P\{2 < X \le 4\}$.

Solution. (a) $P\{X < 3\} = \lim_n P\left\{X \le 3 - \dfrac{1}{n}\right\} = \lim_n F\left(3 - \dfrac{1}{n}\right) = \dfrac{11}{12}$

(b)

$$\begin{aligned} P\{X = 1\} &= P\{X \le 1\} - P\{X < 1\} \\ &= F(1) - \lim_n F\left(1 - \frac{1}{n}\right) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6} \end{aligned}$$

(c)

$$\begin{aligned} P\left\{X > \frac{1}{2}\right\} &= 1 - P\left\{X \le \frac{1}{2}\right\} \\ &= 1 - F\left(\frac{1}{2}\right) = \frac{3}{4} \end{aligned}$$

(d)

$$\begin{aligned} P\{2 < X \le 4\} &= F(4) - F(2) \\ &= \frac{1}{12} \end{aligned}$$

■

SUMMARY

~~SUMMARY~~

A real-valued function defined on the outcome of a probability experiment is called a *random variable*.

If X is a random variable, then the function $F(x)$ defined by

$$F(x) = P\{X \leq x\}$$

is called the *distribution function* of X. All probabilities concerning X can be stated in terms of F.

A random variable whose set of possible values is either finite or countably infinite is called *discrete*. If X is a discrete random variable, then the function

$$p(x) = P\{X = x\}$$

is called the *probability mass function* of X. Also, the quantity $E[X]$ defined by

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

is called the *expected value* of X. $E[X]$ is also commonly called the *mean* or the *expectation* of X.

A useful identity states that, for a function g,

$$E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$$

The *variance* of a random variable X, denoted by $\mathrm{Var}(X)$, is defined by

$$\mathrm{Var}(X) = E[(X - E[X])^2]$$

The variance, which is equal to the expected square of the difference between X and its expected value, is a measure of the spread of the possible values of X. A useful identity is

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

The quantity $\sqrt{\mathrm{Var}(X)}$ is called the *standard deviation* of X.

We now note some common types of discrete random variables. The random variable X whose probability mass function is given by

$$p(i) = \binom{n}{i} p^i(1 - p)^{n-i} \quad i = 0, \ldots, n$$

is said to be a binomial random variable with parameters n and p. Such a random variable can be interpreted as being the number of successes that occur when n independent trials, each of which results in a success with probability p, are performed. Its mean and variance are given by

$$E[X] = np \quad \mathrm{Var}(X) = np(1 - p)$$

The random variable X whose probability mass function is given by

$$p(i) = \frac{e^{-\lambda}\lambda^i}{i!} \quad i \geq 0$$

is said to be a *Poisson* random variable with parameter λ. If a large number of (approximately) independent trials are performed, each having a small probability of being successful, then the number of successful trials that result will have a distribution which is approximately that of a Poisson random variable. The mean and variance of a Poisson random variable are both equal to its parameter λ. That is,

$$E[X] = \mathrm{Var}(X) = \lambda$$

The random variable X whose probability mass function is given by

$$p(i) = p(1 - p)^{i-1} \quad i = 1, 2, \ldots$$

is said to be a *geometric* random variable with parameter p. Such a random variable represents the trial number of the first success when each trial is independently a success with probability p. Its mean and variance are given by

$$E[X] = \frac{1}{p} \quad \mathrm{Var}(X) = \frac{1 - p}{p^2}$$

The random variable X whose probability mass function is given by

$$p(i) = \binom{i-1}{r-1} p^r (1 - p)^{i-r} \quad i \geq r$$

is said to be a *negative binomial* random variable with parameters r and p. Such a random variable represents the trial number of the rth success when each trial is independently a success with probability p. Its mean and variance are given by

$$E[X] = \frac{r}{p} \quad \mathrm{Var}(X) = \frac{r(1 - p)}{p^2}$$

A *hypergeometric* random variable X with parameters n, N, and m represents the number of white balls selected when n balls are randomly chosen from an urn that contains N balls of which m are white. The probability mass function of this random variable is given by

$$p(i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \quad i = 0, \ldots, m$$

With $p = m/N$, its mean and variance are

$$E[X] = np \quad \mathrm{Var}(X) = \frac{N - n}{N - 1} np(1 - p)$$

An important property of the expected value is that the expected value of a sum of random variables is equal to the sum of their expected values. That is,

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$